



# Messages From the Past: New Insights in Plant Lectin Evolution

Sofie Van Holle and Els J. M. Van Damme\*

Laboratory of Biochemistry and Glycobiology, Department of Biotechnology, Ghent University, Ghent, Belgium

## OPEN ACCESS

### Edited by:

Stefan A. Rensing,  
University of Marburg, Germany

### Reviewed by:

Andrew Charles Cuming,  
University of Leeds, United Kingdom  
Hervé Canut,  
UMR5546 Laboratoire de Recherche  
en Sciences Vegetales (LRSV), France

### \*Correspondence:

Els J. M. Van Damme  
ElsJM.VanDamme@UGent.be

### Specialty section:

This article was submitted to  
Plant Development and EvoDevo,  
a section of the journal  
Frontiers in Plant Science

**Received:** 14 July 2018

**Accepted:** 10 January 2019

**Published:** 29 January 2019

### Citation:

Van Holle S and Van Damme EJM  
(2019) Messages From the Past: New  
Insights in Plant Lectin Evolution.  
Front. Plant Sci. 10:36.  
doi: 10.3389/fpls.2019.00036

Lectins are a large and diverse class of proteins, found in all kingdoms of life. Plants are known to express different types of carbohydrate-binding proteins, each containing at least one particular lectin domain which enables them to specifically recognize and bind carbohydrate structures. The group of plant lectins is heterogeneous in terms of structure, biological activity and function. Lectins control various aspects of plant development and defense. Some lectins facilitate recognition of exogenous danger signals or play a role in endogenous signaling pathways, while others are considered as storage proteins or involved in symbiotic relationships. In this study, we revisit the origin of the different plant lectin families in view of the recently reshaped tree of life. Due to new genomic sampling of previously unknown microbial lineages, the tree of life has expanded and was reshaped multiple times. In addition, more plant genomes especially from basal Phragmoplastophyta, bryophytes, and Salviniales (e.g., *Chara braunii*, *Marchantia polymorpha*, *Physcomitrella patens*, *Azolla filiculoides*, and *Salvinia cucullata*) have been analyzed, and annotated genome sequences have become accessible. We searched 38 plant genome sequences including core eudicots, monocots, gymnosperms, fern, lycophytes, bryophytes, charophytes, chlorophytes, glaucophytes, and rhodophytes for lectin motifs, performed an extensive comparative analysis of lectin domain architectures, and determined the phylogenetic and evolutionary history of lectins in the plant lineage. In conclusion, we describe the conservation of particular domains in plant lectin sequences obtained from algae to higher plants. The strong conservation of several lectin motifs highlights their significance for plants.

**Keywords:** lectin, gene family evolution, lower plants, protein domain, evolutionary diversity

## INTRODUCTION

After the discovery of the Archaea, a model was proposed that divides cellular life into three evolutionary domains: “Eukarya,” “Bacteria,” and “Archaea” (Woese et al., 1990). In this three-domain tree of life, Archaea and Eukarya are sister groups that share a common ancestor. Over the years, the evolutionary relationships between Archaea and Eukarya have been the subject of long-lasting debates (Williams et al., 2013; Forterre, 2015; Hug et al., 2016). Recent comprehension of novel archaeal superphyla through metagenomic analyses and advances in molecular phylogenetics provided a novel view on the origin and early evolution of eukaryotes. Today, the two-domain topology is generally accepted, with Bacteria and Archaea being the two primary branches, in which eukaryotes have emerged from within the Archaea. Taking into account the most recent

phylogenomic analyses, eukaryotes most probably originate from within the Asgard (being the closest prokaryotic relatives of eukaryotes) and not the ‘TACK’ (which groups Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota) superphylum within the Archaea (Eme et al., 2017; Zaremba-Niedzwiedzka et al., 2017). In addition, phylotranscriptomic data show that approximately 450–500 million years ago, land plants evolved from a streptophyte algae lineage (Zygnematophyceae) (Wickett et al., 2014). The transition from unicellular and filamentous algae to modern land plants required distinctive adaptations/exaptations to the terrestrial environment including three-dimensional growth, sporophyte dominance, development of vasculature and desiccation-tolerant seeds (Harrison, 2017; Rensing, 2018; de Vries and Archibald, 2018).

Lectins are a group of diverse proteins that occur ubiquitously in nature and share the ability to recognize and bind specific carbohydrate structures. Plants lectins are mainly involved in plant immunity and symbiosis, but roles in plant development have also been attributed to particular lectins (reviewed by Van Holle and Van Damme, 2018). For a long time, most research aimed at the biochemical and functional characterization of plant lectins, while their relevance in the colonization of land by plants and in the evolution of angiosperms was neglected. Today, various studies report on the abundance of lectin genes in modern plant models and homologs of plant lectins have also been reported outside the plant kingdom (Naganuma et al., 2014; Wong et al., 2014). A recent study on lectin sequences in model species (*Arabidopsis*, rice, soybean, and cucumber) points to a dynamic evolution of these protein families (Van Holle et al., 2017). Unfortunately, only angiosperm genomes were included in these analyses, which makes it difficult to reconstruct how plant lectins diverged from their common ancestor. In 2018, the first genomic data from ferns (*Azolla filiculoides* and *Salvinia cucullata*), a close sister group to angiosperms, was published (Li et al., 2018). Furthermore, a new chromosome-scale assembly of the *Physcomitrella patens* genome, a model for the mosses, was first released in 2017 (Lang et al., 2018). The genome sequences of *Marchantia polymorpha*, a model species for the liverwort lineage, is believed to represent the last common ancestor of extant land plants the best because of its low genetic redundancy (Bowman et al., 2017). However, the true bryophyte topology is still enigmatic and the relevance of the liverwort *Marchantia polymorpha* as a model for the earliest land plants is heavily disputed (Puttick et al., 2018; Rensing, 2018). The genome of *Chara braunii* (Nishiyama et al., 2018) and of *Klebsormidium nitens* (Hori et al., 2014) represent the Charophyceae and Klebsormidiophyceae, charophycean algae that share a common ancestor with land plants. In addition to these key genomes of the streptophyte lineage, the study of genome sequences from Chlorophyta (including prasinophytes and core chlorophytes), the freshwater microscopic algae *Cyanophora paradoxa* (Price et al., 2012) and rhodophytes (*Cyanidioschyzon merolae* and *Porphyra umbilicalis*) (Matsuzaki et al., 2004; Brawley et al., 2017) can further refine the divergence of the plant lectin family and their establishment during land plant evolution.

Since substantial progress has been made recently in resolving the placement of eukaryotes within the Archaea, the primary

focus of our study relates to the origin of plant lectins in the tree of life. We attempted to reconstruct the evolutionary origins of the plant lectin families. Our data highlight that some families are a eukaryotic innovation, while others are descendants of ancient protein families as they are also found in prokaryotes. We also considered the domain architectures and diversification of specific lectin families, with emphasis on the similarities/differences between land plant lineages and in lineages that are sister to land plants.

## MATERIALS AND METHODS

### Assembly of Dataset of Plant Lectin Homologs in the Tree of Life

In this study, eleven plant lectin families (Van Damme et al., 1998) [represented by *Agaricus bisporus* agglutinin, amaranthin, cyanovirin, *Euonymus*-related lectin (EUL), *Galanthus nivalis* lectin (GNA), hevein, jacalin, legume lectin, LysM: lysin motif, *Nicotiana tabacum* agglutinin (Nictaba) and ricin B], together with the malectin family -initially reported in Metazoa (Schallus et al., 2008)- have been considered. The availability of a unique Pfam and Interpro identifier for each lectin family facilitated a straightforward search for lectin sequences in a large dataset. In total, 38 plant genomes were selected, representing diverse clades of Archaeplastida, including 15 core eudicots (*Arabidopsis thaliana*, *Capsella rubella*, *Brassica rapa*, *Theobroma cacao*, *Citrus clementina*, *Populus trichocarpa*, *Linum usitatissimum*, *Ricinus communis*, *Malus domestica*, *Cucumis sativus*, *Glycine max*, *Medicago truncatula*, *Vitis vinifera*, *Solanum lycopersicum*, and *Amaranthus hypochondriacus*), five monocots (*Zea mays*, *Sorghum bicolor*, *Brachypodium distachyon*, *Oryza sativa*, and *Musa acuminata*), one basal angiosperm (*Amborella trichopoda*), two gymnosperms (*Ginkgo biloba* and *Picea abies*), two Polypodiopsida (*Azolla filiculoides* and *Salvinia cucullata*), one lycophyte (*Selaginella moellendorffii*), three bryophytes (*Physcomitrella patens*, *Sphagnum fallax* and *Marchantia polymorpha*), one Charophyceae (*Chara braunii*), one Klebsormidiophyceae (*Klebsormidium nitens*), two core chlorophytes (*Chlorella* NC64A and *Chlamydomonas reinhardtii*), two prasinophytes (*Ostreococcus lucimarinus* and *Micromonas* sp. RCC299), one glaucophyte (*Cyanophora paradoxa*) and two rhodophytes (*Cyanidioschyzon merolae* and *Porphyra umbilicalis*). Exploration of the plant comparative genomics platform PLAZA 4.0<sup>1</sup> (Van Bel et al., 2018) employing the Interpro identifier revealed the abundance of lectin domains. Following Interpro identifiers were used: IPR009960 (*Agaricus bisporus* agglutinin), IPR008998 (amaranthin), IPR011058 (cyanovirin), IPR001480 (GNA), IPR001002 (hevein), IPR001229 (jacalin), IPR001220 (legume lectin), IPR018392 (LysM), IPR024788 and IPR021720 (malectin), IPR025886 (Nictaba), and IPR000772 (ricin B). The domain architecture of the retrieved sequences was analyzed using the Interpro, incorporated in the PLAZA 4.0 toolbox. Because EUL and ricin B lectin domains are characterized by the same identifier, the extent of the EUL family

<sup>1</sup> [https://bioinformatics.psb.ugent.be/plaza/versions/plaza\\_v4\\_dicots/](https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_dicots/)

was estimated based on the results from Fouquaert et al. (2009a) and De Schutter et al. (2017). Additionally, Gymno PLAZA 1.0<sup>2</sup> (Proost et al., 2015) and Phytozome v12.1.5<sup>3</sup> (Goodstein et al., 2012) were employed to investigate the distribution of lectin sequences in *Ginkgo biloba* and *Musa acuminata*, *Linum usitatissimum*, *Sorghum bicolor*, *Brachypodium distachyon*, *Sphagnum fallax*, *Ostreococcus lucimarinus*, *Micromonas* sp. RCC299, and *Porphyra umbilicalis* using a similar approach. The Interpro or Pfam identifier was used to search for sequences accommodating the lectin motifs and the amino acid sequences were downloaded. Next, the sequences were subjected to analysis with Interpro 71.0 at <https://www.ebi.ac.uk/interpro/> (Mitchell et al., 2015) to explore the domain architecture. Putative lectin sequences from *Cyanidioschyzon merolae*, *Cyanophora paradoxa*, *Chlorella* NC64A, *Klebsormidium nitens*, *Chara braunii*, *Azolla filiculoides*, and *Salvinia cucullata* were identified through BLASTp searches online<sup>4, 5, 6, 7, 8, 9</sup> using a representative query for each of the lectin families. Using the following settings, expect value: 10 and substitution matrix: BLOSUM62, the best hit was used for a consecutive BLASTp search to obtain all possible lectin sequences. Next, the presence of lectin domains (and other protein domains) was verified with Interpro 71.0 at <https://www.ebi.ac.uk/interpro/> (Mitchell et al., 2015). **Supplementary Table 1** compiles the obtained results of this large genome-wide search for lectin motifs. The Interpro database (71.0) was used to estimate the occurrence of lectin domains in different lineages in the tree of life. The results are compiled in **Figure 1** and represent estimates of the size of the different lectin families.

## Evolutionary Expansion of Plant Lectin Genes

To understand the evolution of the hevein, jacalin and Nictaba gene family, reconciliation trees were generated using Notung 2.9 (Stolzer et al., 2012). Trees were rooted in Notung and in the reconciliation mode, the edge weight threshold was set to 0.0005, the loss cost to 1, the duplication to 1.5 and the co-divergence cost to 0. The species tree was constructed in NCBI Taxonomy Common Tree<sup>10</sup> and contained the following species: *Amaranthus hypochondriacus*, *Amborella trichopoda*, *Arabidopsis thaliana*, *Azolla filiculoides*, *Brassica rapa*, *Capsella rubella*, *Chara braunii*, *Chlamydomonas reinhardtii*, *Chlorella* NC64A, *Cucumis sativus*, *Cyanidioschyzon merolae*, *Cyanophora paradoxa*, *Ginkgo biloba*, *Glycine max*, *Klebsormidium nitens*, *Linum usitatissimum*, *Malus domestica*, *Marchantia polymorpha*, *Micromonas* sp. RCC299, *Musa acuminata*, *Oryza sativa*, *Picea*

*abies*, *Physcomitrella patens*, *Populus trichocarpa*, *Porphyra umbilicalis*, *Salvinia cucullata*, *Selaginella moellendorffii*, *Solanum lycopersicum*, *Sorghum bicolor*, *Sphagnum fallax*, *Theobroma cacao*, *Vitis vinifera*. Amino acid sequences for hevein, jacalin and Nictaba homologs present in these species were downloaded from the Phytozome 12.1 database (see footnote 3), PLAZA 4.0 (see footnote 1), Gymno PLAZA 1.0<sup>2</sup> (see footnotes 2, 6, 7, 8, 9). Multiple sequence alignments were performed online with MAFFT version 7<sup>11</sup> using default settings (Katoh et al., 2017). Subsequently, the aligned sequences were subjected to trimAl v.3 (Capella-Gutiérrez et al., 2009) to trim the alignments to the gap threshold of “gt 0.6” and remove all columns with gaps in more than 40% of the sequences. Next, maximum likelihood gene trees were constructed with IQ-TREE 1.3.1 [automated selection of the substitution model including FreeRate heterogeneity, ultrafast bootstrap approximation (UFBoot) with 1,000 bootstrap alignments] (Nguyen et al., 2015; Kalyanamoorthy et al., 2017; Hoang et al., 2018). Whole genome duplication and triplication events as described by Li et al. (2016, 2018) and Lang et al. (2018) were added to the gene trees reconciled with the species tree. Figtree v. 1.4.3 was used to visualize and modify the phylogenetic trees<sup>12</sup>.

## Sequence Motif Analysis

Sequence motif analysis was performed online with MEME suite 5.0.0<sup>13</sup> (Bailey et al., 2009). Protein datasets were mined for conserved motifs within the lectin domain sequences that were identified in the 38 plant species described above. Parameters were set as follows: classic mode, window size of 6–50. The distribution of selected significant motifs was analyzed across sequences and species.

## RESULTS AND DISCUSSION

### Evidence for Plant Lectin Domains in Bacteria and Archaea

In the plant kingdom, several unique lectin families have been reported and each of them is defined by a characteristic carbohydrate-recognition or lectin domain (Van Damme et al., 2008). Taking advantage of the wealth of available sequenced genomes, we mined the predicted proteomes of all species available in the Interpro 71.0 database for plant lectin motifs. As shown in **Figure 1**, the occurrence of most plant lectin domains is not restricted to the plant kingdom. While the distribution of the amarantins and the EUL family is limited to plants, all other lectin domains are also present in other lineages of the tree of life. However, large differences in the number of sequences within one particular family are observed between the different lineages (**Figure 1**). Furthermore, the discrepancy between lectin families points to distinct evolutionary paths. The malectin family is represented by two Interpro identifiers: the “Malectin domain” (IPR021720) and the “Malectin-like domain” (IPR024788). In

<sup>2</sup><https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/>

<sup>3</sup><https://phytozome.jgi.doe.gov>

<sup>4</sup><http://merolae.biol.s.u-tokyo.ac.jp/blast/blast.html>

<sup>5</sup><http://cyanophora.rutgers.edu/cyanophora/>

<sup>6</sup>[https://genome.jgi.doe.gov/pages/blast-query.jsf?db=ChlNC64A\\_1](https://genome.jgi.doe.gov/pages/blast-query.jsf?db=ChlNC64A_1)

<sup>7</sup>[http://www.plantmorphogenesis.bio.titech.ac.jp/~algae\\_genome\\_project/klebsormidium/klebsormidium\\_blast.html](http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/klebsormidium_blast.html)

<sup>8</sup>[http://bioinformatics.psb.ugent.be/blast/moderated/?project=orcae\\_Chbra](http://bioinformatics.psb.ugent.be/blast/moderated/?project=orcae_Chbra)

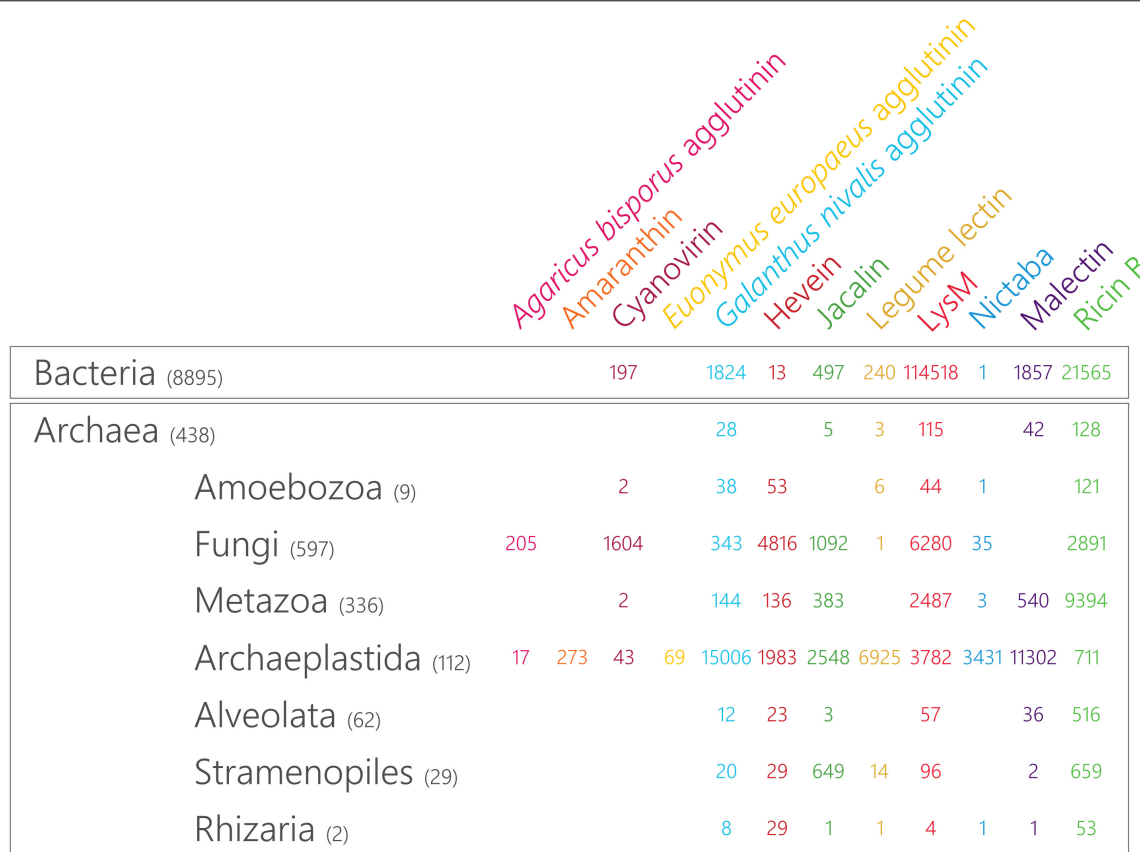
<sup>9</sup><https://www.fernbase.org/>

<sup>10</sup><https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>

<sup>11</sup><https://mafft.cbrc.jp/alignment/server/index.html>

<sup>12</sup><http://tree.bio.ed.ac.uk/software/figtree>

<sup>13</sup><http://meme-suite.org/>



**FIGURE 1 |** Taxonomic distribution and magnitude of plant lectin domains in the tree of life. The size of the lectin domain families is retrieved from Interpro 71.0 and the number of included reference proteomes in each of the lineages is indicated between brackets.

**Figure 1**, the combined number of sequences for both identifiers is shown. It should be mentioned that the “Malectin-like domain” (IPR024788) could only be retrieved in Archaeplastida (with 7,771 hits) and in none of the other lineages.

## Not All Plant Lectin Domains Originate From Within the Archaeplastida

To reconstruct the evolutionary paths for the different plant lectin domains, representative genomes of the most important lineages within Archaeplastida (including core eudicots, monocots, one basal angiosperm, two gymnosperms, two ferns, one lycophyte, three bryophytes, one Charophyceae, one Klebsormidiophyceae, two core chlorophytes, two prasinophytes, one glaucophyte and two rhodophytes) were screened for the presence of plant lectin domains (**Figure 2**). It is clear that the distribution of the lectin motifs is variable, pointing toward a different evolutionary origin for different lectin domains.

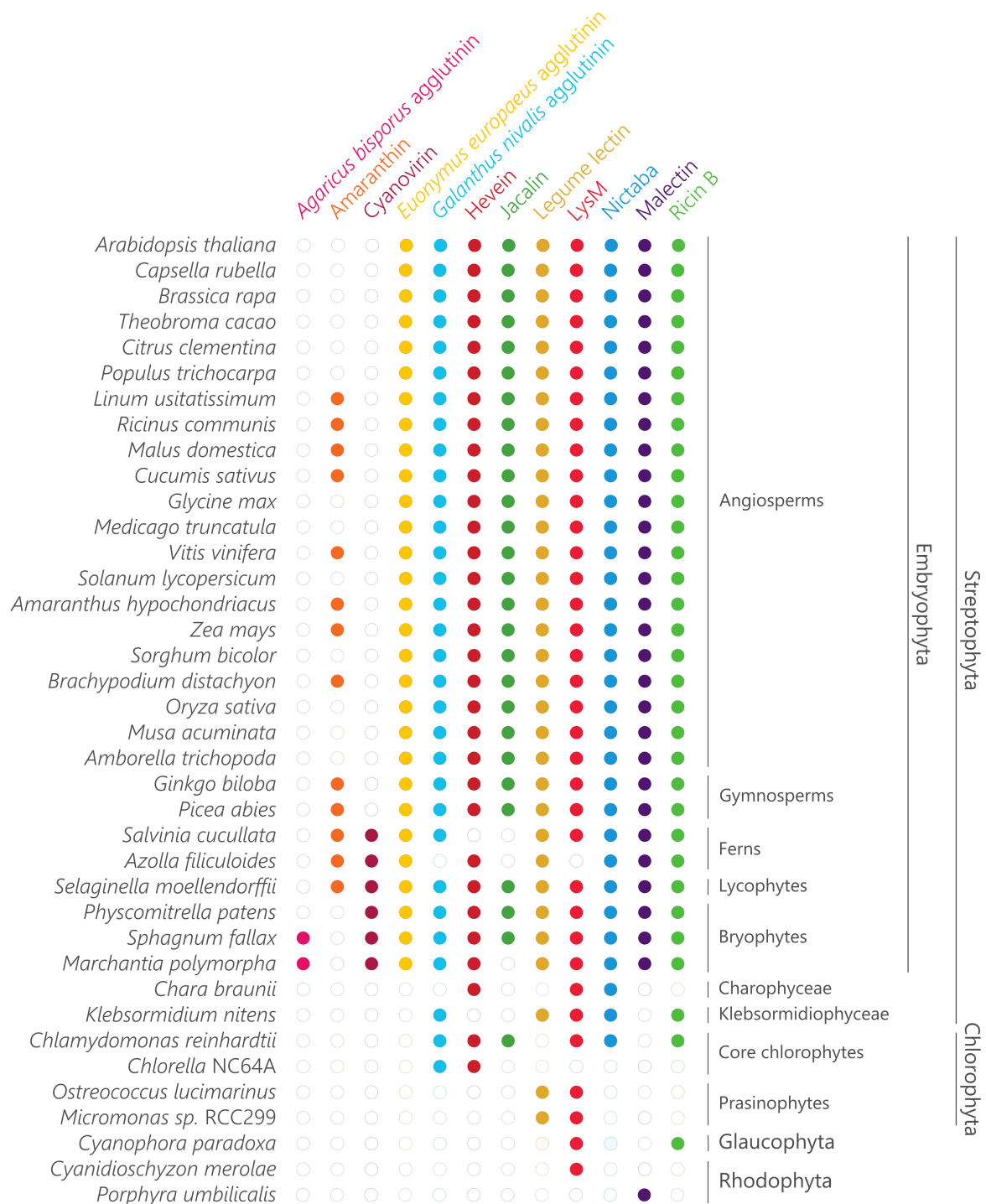
The *Agaricus bisporus* agglutinin originates from fungi and data shown in **Figure 1** confirm that this lineage encompasses the highest number of *Agaricus bisporus* agglutinin homologs. In plants, the presence of this lectin domain is restricted to the bryophytes (**Figure 2**). Homologs of the *Agaricus bisporus* agglutinin were only retrieved from the genomes of the liverwort

*Marchantia polymorpha* and the bog moss *Sphagnum fallax*. This is the first record of a homolog for the *Agaricus bisporus* agglutinin in mosses, as functional lectins of this family have only been described in fungi and *Marchantia polymorpha* (Peumans et al., 2007; Bovi et al., 2011). It can be assumed that this lectin domain arose in a fungal ancestor (**Figure 3**) and that horizontal gene transfer, possibly through endosymbionts, is responsible for its confined existence in Archaeplastida.

A similar story holds for the cyanovirin lectin domain. This lectin domain is mostly present in fungi (**Figure 1**) and to a lesser degree in Bacteria, Amoebozoa, Metazoa, and plants. The rather limited distribution in Bacteria and Eukarya points to multiple independent horizontal gene transfers between fungi and Bacteria, and/or between fungi and an ancestor of Embryophyta (**Figure 3**) as suggested earlier (Percudani et al., 2005). Clearly, the cyanovirin domain was purged during the evolution of gymnosperms and angiosperms (**Figure 2**).

The occurrence of the amaranthin domain is limited to vascular plants (lycophytes, ferns, gymnosperms and angiosperms) and scattered over different families but is certainly not ubiquitous (**Figure 2**). This taxonomic distribution pattern is very mystifying which makes it difficult to reveal the exact phylogeny, but suggests an origin within the vascular plant lineage (**Figure 3**). This is in line with a recent study, in which



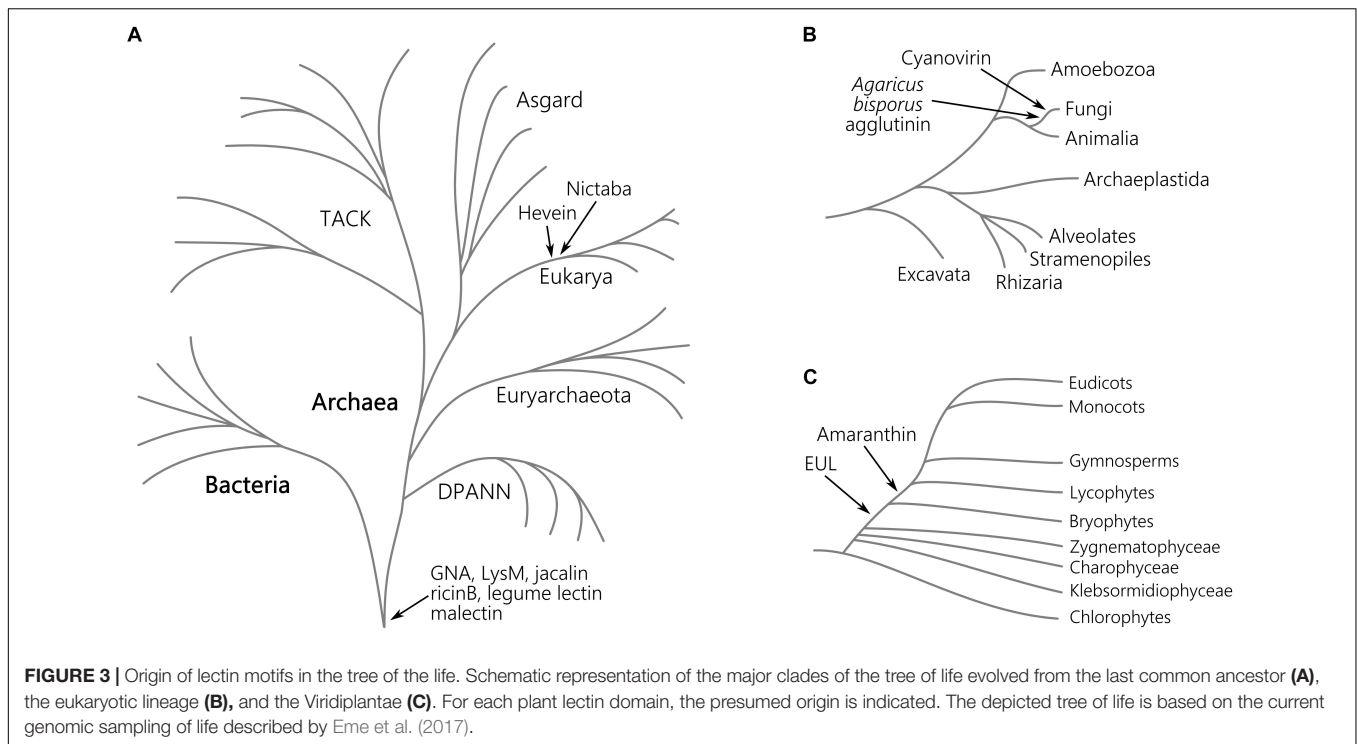


**FIGURE 2 |** Sampling of plant lectin domains from the plant lineage. Coulson plot demonstrating presence or absence of genes encoding plant lectins across a range of Archaeplastida. Filled circles indicate genes identified with high confidence.

amaranthin sequences were identified in 33 plant genomes. Here, a similar complex distribution pattern was also observed (Dang et al., 2017).

Similar to the amarantins, the EUL family represents a true plant lectin family. EUL homologs are found in land

plants (Embryophyta) including the bryophyte lineage, but unlike the amarantins, are omnipresent. Presumably, this protein domain arose in the last common ancestor of the Embryophyta (Figure 3) and remained part of the lectin collection during the development of modern land plants. The results validate



an earlier study in which the complete genome sequence of *Marchantia polymorpha* was not yet available (Fouquaert et al., 2009a). A striking correlation was observed between the origin of the EUL family and the occurrence of stomata. Ancient types of stomata are described in members of the bryophyte lineage (Chater et al., 2017) while homologs of ArathEULS3, a lectin involved in stomatal closure (Van Hove et al., 2015), originate from the same lineage. Deciphering the function of the EUL homologs and other lectins in these extant plant species will bring clarity into their ancestral role, furthering their evolutionary history and explain how they evolved to a diversified group of proteins in higher plants. This could help us to answer the question whether the evolution of the EUL family occurred in parallel with the stomatal development during terrestrial transition of plants, and will be important to elucidate their function.

The distribution of GNA, LysM, jacalin, ricin B, legume lectin and malectin domains in all lineages of the tree of life (Figure 1) proposes an origin in the last universal common ancestor of Bacteria and Archaea (Figure 3). The GNA, jacalin, malectin and legume lectin sequences are more prevalent in plants while LysM and ricin B domains are most abundant in Bacteria (Figure 1). Malectin domains could only be retrieved from embryophyte genomes and the rhodophyte *Porphyra umbilicalis*, while GNA, LysM, ricin B and legume lectins homologs appear to be ubiquitous in Viridiplantae, including Chlorophyta and Streptophyta (Figure 2). In the glaucophyte *Cyanophora paradoxa*, only LysM and ricin B lectin domains could be retrieved. LysM lectin domains were also identified in one of the two rhodophyte species under study, *Cyanidioschyzon merolae*. Especially for the jacalin and LysM lectin family, several

studies have already reported on their widespread distribution (Nagata et al., 2005; Buist et al., 2008; Zhang et al., 2009; Kanagawa et al., 2014; Naganuma et al., 2014; Akcapinar et al., 2015).

Judging from Figure 1, hevein and Nictaba-related lectins are a eukaryotic innovation. Because the hevein domain is shared by the most significant eukaryotic lineages (Stramenopiles, Rhizaria, Alveolata, Amoebozoa, Archaeplastida, Fungi, and Metazoa), it must have arisen in their last common ancestor, and evolved independently after the lineages split (Figure 3). Our data show that the number of hevein sequences varies considerably with more than 1,900 homologs in fungi and Archaeplastida (Viridiplantae, Glaucophyta, and Rhodophyta), 136 sequences in Metazoa (animals) and less than 54 homologs in the other clades. The hevein domain has also been reported in some nematode species (Bauters et al., 2017). The hevein domain is absent from Archaea, but 13 homologs were mined from Bacteria, in particular in Burkholderiales and Brenneria, containing mostly plant pathogenic bacteria (Ura et al., 2006; Young and Park, 2007; Maes et al., 2009; Ham et al., 2011; Lee et al., 2016). The Nictaba domain on the other hand, is more confined to Archaeplastida and fungi since three hits were found in Metazoa and only one hit in Rhizaria, Amoebozoa, and Bacteria. Differential loss of homologous genes in the genomes of Amoebozoa, Rhizaria, and Metazoa, rather than multiple independent horizontal gene transfers possibly account for the complex pattern of Nictaba sequences in Eukaryota. In the green lineage, both Nictaba and hevein sequences are widespread and present in land plants and core chlorophytes. In Klebsormidiophyceae, only Nictaba sequences are found while Charophyceae contain hevein homologs in addition to Nictaba lectins (Figure 2).

## Lectin Homologs Are Variably Maintained Across a Broad Range of Plant Lineages

Plant lectin genes were exploited in key genomes of significant lineages in Archaeplastida. The comparative analysis of the lectin sequences retrieved from the rhodophytes *Cyanidioschyzon merolae* and *Porphyra umbilicalis*; the glaucophyte *Cyanophora paradoxa*; chlorophytes *Micromonas* sp. RCC299, *Chlorella* NC64A, *Chlamydomonas reinhardtii*; Klebsormidiophyceae *Klebsormidium nitens*; Charophyceae *Chara braunii*; bryophytes *Marchantia polymorpha*, *Physcomitrella patens*; Polypodiopsida *Azolla filiculoides*, *Salvinia cucullata* and gymnosperms *Picea abies* and *Ginkgo biloba*; revealed a large discrepancy in the organization and distribution of the lectin families (Supplementary Tables 1, 2 and Figure 2). Representatives of the *Agaricus bisporus* agglutinin are confined to the genomes of *Marchantia polymorpha* and *Sphagnum fallax* while amarantins are not yet present in this plant lineage. The earliest records of amarantin homologs are found in lycophytes. EUL homologs are found in bryophytes and vascular plants and represent a rather small family, except in the *Ginkgo biloba* genome. Similarly, the cyanovirin family present in bryophytes, lycophytes and ferns only represents a small fraction of the lectin collection in these species. Remarkably, it is the second largest lectin family (22.4%) in *Salvinia cucullata*. GNA homologs appear as a large fraction of the total number of lectin genes in *Marchantia polymorpha*, *Physcomitrella patens* and *Picea abies* but are a rather small family in *Chlamydomonas reinhardtii*, *Klebsormidium nitens*, and *Ginkgo biloba*. The size of the hevein family is very much dependent on the plant species, analogously to GNA homologs. Lectin homologs belonging to the jacalin and LysM family represent more than 70% of the total number of lectin genes in the chlorophyte *Chlamydomonas reinhardtii*, whereas no jacalin-related lectins could be retrieved from most other plant species. Legume lectin sequences account for the largest lectin families in *Marchantia polymorpha*, *Azolla filiculoides*, *Salvinia cucullata*, and *Picea abies*. Only one or two lectin motifs are identified in *Cyanophora paradoxa* (LysM), *Micromonas* sp. RCC299 (legume lectin, LysM), *Chlorella* NC64A (GNA, hevein) and *Cyanidioschyzon merolae* (LysM) while only malectins were retrieved from the genome of *Porphyra umbilicalis*. Clearly, there is no evidence for widespread or abundant lectin motifs in prasinophytes, glaucophytes, and rhodophytes.

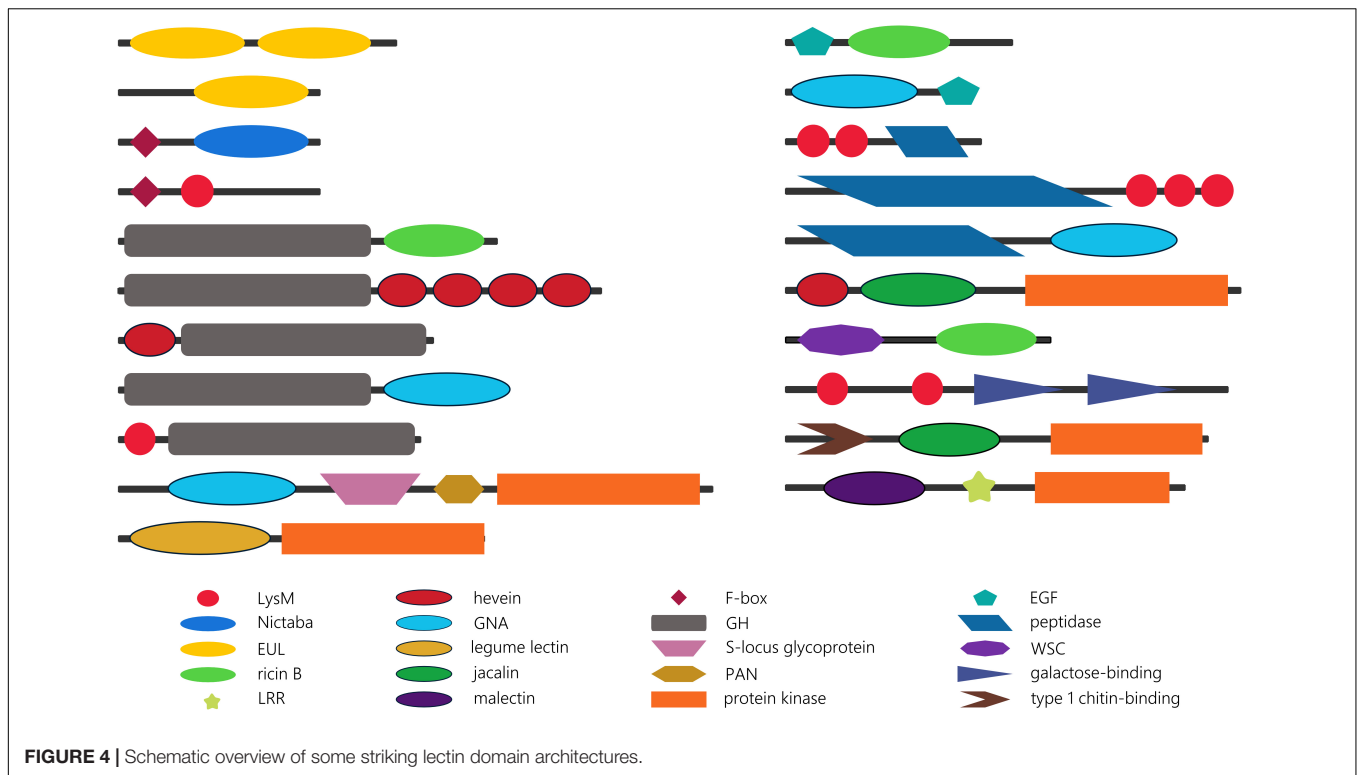
## Diversification of Domain Arrangements in Higher Plant Lineages

To gain more insight into their evolutionary history, the domain organization of the lectin sequences was investigated. The presence of multi-domain proteins in the genomes of all kingdoms of life has been reported before (Ekman et al., 2005). As a result of their more complex genome and biology, higher eukaryotes are considered to display a larger collection of multi-domain proteins (Ekman et al., 2007). Supplementary Table 3 summarizes the domain architectures of plant lectin sequences in 14 rhodophytes, glaucophytes, chlorophytes, Klebsormidiophyceae, Charophyceae, bryophytes, lycophytes,

Polypodiopsida and gymnosperms, and the preservation of these domain architectures in four core model angiosperms (*Arabidopsis thaliana*, *Glycine max*, *Cucumis sativus*, and *Oryza sativa*). A comprehensive analysis of all the protein domains associated with plant lectin domains in the plant species under study (*Cyanidioschyzon merolae*, *Porphyra umbilicalis*, *Cyanophora paradoxa*, *Micromonas* sp. RCC299, *Chlorella* NC64A, *Chlamydomonas reinhardtii*, *Klebsormidium nitens*, *Chara braunii*, *Marchantia polymorpha*, *Physcomitrella patens*, *Azolla filiculoides*, *Salvinia cucullata*, *Picea abies*, and *Ginkgo biloba*) yielded an extensive list of protein domain arrangements (Supplementary Table 3). The description of each protein domain and lectin domain combination is beyond the scope of this manuscript. Below we describe some striking observations and interesting domain combinations (Figure 4).

The EUL family groups all sequences with an *Euonymus*-related lectin domain. Though the EUL lectin domain can be preceded or followed by sequences longer than 100 amino acids, no protein domains other than the lectin domain are recognized in the EUL sequences. This characteristic is unique for the EUL family. Two types of EUL domain architectures have been described: proteins consisting of two tandem arrayed EUL domains and single EUL domain proteins (Fouquaert et al., 2009a). Both single and double EUL domain proteins are present in the genomes of *Marchantia polymorpha* and *Physcomitrella patens* (Supplementary Table 3). This trait is shared with monocot lineages, while genomes from dicot species exclusively harbor single EUL domain architectures (De Schutter et al., 2017). Similar to the core eudicot genomes, single EUL domain proteins were identified in the genomes from the Polypodiopsida and the gymnosperms under study. Until now, it remains unclear why the Eudicotyledones did not maintain the double EUL domain architecture in their genome.

Unlike the EUL lectin family, sequences from all other plant lectin families are composed of lectin domains linked with a variety of other annotated protein domains. Though most lectin sequences are composed of two protein domains, lectin sequences with up to six different protein domains have been reported (Van Holle et al., 2017). Analysis of the identified domain architectures in the lectin sequences retrieved from the genomes that are sister to vascular plants, land plants, Streptophyta and Viridiplantae under study, revealed some remarkable similarities with the domain organization of lectin sequences in higher plants. Similar to the model species *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max* and *Cucumis sativus*; glycoside hydrolase (GH), protein kinase or F-box domains are often found in combination with lectin domains in sisters of Streptophyta and vascular plants. It could be postulated that most of these domain architectures arose in a common ancestor of Streptophyta and/or chlorophytes, and that have been retained during evolution. However, there are a number of anomalies. The combination of an F-box domain with a Nictaba or LysM domain is shared by all Viridiplantae, while the F-box/jacalin combination is only present in higher plants (Van Holle et al., 2017). Similarly protein architectures involving GH and hevein or ricin domains are ubiquitous in Viridiplantae while the association of a GH domain and a GNA or LysM domain in the genomes under



study is restricted to bryophytes. Moreover, the combination of GH and hevein domains is determinative in terms of the order of the domains, and in terms of the type of GH. In the PLAZA 4.0 database, hevein/GH sequences are defined as Embryophyta-specific, while GH/hevein domain organization is Chlorophyta-specific. Furthermore, the hevein/GH18 domain combination is only present in bryophytes; in contrast to the hevein/GH19 combination which is shared by Embryophyta (Supplementary Table 3). Previous studies reported on the expansion of the receptor-like protein kinases in the plant lineage (Lehti-Shiu and Shiu, 2012; Xing et al., 2013) and suggest that receptor-like protein kinases originated within the Streptophyta lineage, with a significant increase in gene number in angiosperms. Embryophyta are the earliest lineage in which lectin receptor-like protein kinases are found, as illustrated in the GNA family. The GNA/S-locus glycoprotein/PAN/protein kinase domain combination is found twice in the genome of *Physcomitrella patens* but up to 76 and 86 sequences with the same architecture are present in the genome of *Oryza sativa* and *Glycine max*, respectively. The expansion of legume/protein kinase sequences evolved in a similar way, but there is also a very large set of homologous sequences in *Marchantia polymorpha* (Supplementary Table 3). The first record of a lectin receptor-like protein kinase (LysM-RLK) was reported in the *Chara braunii* genome, and recently confirmed by the work of Nishiyama et al. (2018). No lectin receptor-like protein kinases were retrieved from the chlorophytes nor in more distant lineages (glaucophytes or rhodophytes). However, the combination of the LysM domain and a protein kinase domain is also present in bacteria (e.g., UniProt A0A0N0UXM0; A1ZLP4), suggesting

that these identical domain architectures arose independently in different lineages.

The lectin sequences retrieved from sisters of angiosperms and from algae revealed some new domain combinations that are only present in rhodophytes, glaucophytes, chlorophytes, charophytes, bryophytes, ferns and/or gymnosperms. Several peptidase domains and the epidermal growth factor (EGF)-like domain are examples of protein domains that are not found in association with lectin domains in angiosperms (Supplementary Table 3). The EGF domain might not be common in the plant kingdom, but is also present in animal lectins. In particular in C-type lectins, the EGF domain is associated with the C-type lectin domain in many different domain arrangements. Some of them include the combination of a C-type lectin domain with multiple EGF domains, sometimes in combination with other protein domains. In vertebrates, C-type lectins have numerous functions, the most important being key players in pathogen sensing and the initiation of immune responses (Mayer et al., 2017; Xia et al., 2018). In general, proteins with EGF domains are predominantly found in a large number of animal protein sequences (Zeng and Harris, 2014). In our analysis, combinations of the EGF-like domain with the GNA and ricin B lectin domains were identified in chlorophytes and/or bryophytes. Thus, it can be postulated that the presence of EGF/lectin domain combinations in these species originates from a eukaryotic lineage, the ancestor of both the plant and animal lineage. The EGF-like domain was originally preserved in chlorophytes and bryophytes, but was subsequently eliminated from the gene set in modern plants.

Peptidase M23 and peptidase C1A domains are associated with LysM domains and the peptidase M11 domain is found in



combination with the GNA domain. These examples illustrate the range of specificities of the peptidase domains. The M11 peptidase is a metalloprotease from *Chlamydomonas reinhardtii* that is involved in cell wall degradation. Next to *Chlamydomonas reinhardtii*, it was also reported in *Volvox carteri* (Kubo et al., 2002). The M23 peptidase has a bacterial origin, similar to the LysM domain to which it is associated. In Archaeplastida, combinations of the M23 peptidase domain and LysM domain were identified in glaucophyta, charophyta, and bryophytes. All these sequences contain multiple LysM domains. In contrast to the M23 peptidases, the C1A peptidases represent mainly a eukaryotic family, with homologs in both the plant and animal kingdom (Santamaría et al., 2014). Although this protein domain is widespread in Viridiplantae, sequences involving a combination of the C1A peptidase and lectin domains have not been retained in vascular plants.

Another striking observation is the unique combination of two different lectin domains (in particular a hevein domain and a jacalin domain, a LysM domain and the fucoselectin tachylectin-4 pentraxin-1 domain, and a LysM domain and two C-type lectin domains) in *Chlamydomonas reinhardtii*. In Nematoda, sequences involving a hevein domain and multiple LysM domains were previously reported (Bauters et al., 2017). Furthermore, domain architectures in which both a lectin domain (ricin B, LysM or jacalin) and at least one other sugar-binding domain (carbohydrate-binding WSC, galactose-binding domain, type 1 chitin-binding domain) have been identified (Supplementary Table 3). In the latter case, the lectin domain and the additional carbohydrate-binding domain most probably display a different carbohydrate-binding specificity. It should be mentioned that further studies at protein level are needed to investigate the functionality of the domains, since the carbohydrate-binding activity of lectin domains cannot be guaranteed based on the presence of a protein sequence.

Regarding the domain arrangement of lectin sequences in basal plant lineages, the multitude of sequences with tandem arrayed lectin domains is noteworthy. Sequences with a two or three LysM domains (in combination with a protein kinase domain) are conserved throughout Archaeplastida. In Arabidopsis and rice, they were identified as part of the plant immune system where they play key roles in the perception and recognition of danger signals. Similar proteins in legumes facilitate symbiotic communication (Zipfel and Oldroyd, 2017; Van Holle and Van Damme, 2018). A sequence composed of four hevein domains was already described in rice (Van Holle et al., 2017), but domain architectures involving more than two hevein domains and additional protein kinase or GH domains appear to be specific to ferns, Chlorophyta or Marchantiaceae.

## Genomic Evolution and Expansion of Nictaba, Jacalin, and Hevein Lectins

Investigation of the expansion of the lectin families during the course of evolution can be linked to specific adaptive speciation events. Three plant lectin families that are present

in both land plants and chlorophytes were selected for detailed analysis. To study the genomic evolution and expansion, the Nictaba, jacalin and hevein gene trees were reconciled with a species tree, including 29 plant genomes (Supplementary Figure 1). The full reconciliation of the Nictaba, jacalin and hevein family trees with the species tree are illustrated in Supplementary Figures 2–4. The Nictaba, jacalin and hevein gene trees are shown in Supplementary Figures 5–7. Supplementary Table 4 summarizes the number of duplications, co-divergences and losses within each of the species, families and ranks.

The Nictaba family evolved through 349 duplications and 314 losses, whereas the jacalin family underwent 287 duplications and 316 losses. In contrast, during the evolution of the hevein family (the smallest in gene size), gene losses were far more abundant (370) than duplication events (216). Whole genome duplication and triplication events have been added to the species tree in Supplementary Figure 1 and are generally believed to play an important role in the expansion of gene families (Soltis et al., 2015; Panchy et al., 2016; Soltis and Soltis, 2016; Van de Peer et al., 2017). Indeed, the two duplication events that are shared by all Brassicaceae resulted in high duplication numbers for all lectin families, yet the *Brassica rapa*-specific whole genome triplication event only contributes to a high number of duplications and losses in the jacalin family. A recent study showed that the *Physcomitrella patens* genome was subjected to two rounds of whole genome duplications. There is evidence that these events are common for mosses while they were not detected in the liverwort (*Marchantia polymorpha*) and hornwort lineages (Lang et al., 2018). None of the lectin genes identified in *Physcomitrella patens* were present in the ancestral karyotype and this is also reflected by the relatively low numbers of duplication and losses for *Physcomitrella patens* (Supplementary Table 4). Overall, the expansion of the jacalin and hevein family in bryophytes is more pronounced compared to the Nictaba family. On the other hand, the latter lectin family displays a larger number of losses in fabids, accompanied by many duplications in all species of this clade. Regarding the evolution of the hevein lectin family, high duplication rates were observed in *Marchantia polymorpha*, *Picea abies*, *Amaranthus hypochondriacus*, *Solanum lycopersicum* and *Populus trichocarpa*. The most important duplication events for the jacalin family are assigned to *Brassica rapa*, *Musa acuminata*, *Oryza sativa*, *Sphagnum fallax* and *Selaginella moellendorffii*.

A large-scale study on gene duplicability across angiosperms revealed that gene duplicability is a non-random process and that most gene families are either primarily single-copy genes or multi-copy genes. Single-copy genes are related to basic cellular functions (organelle function, genome stability maintenance) whereas multi-copy genes are biased toward signaling, transport, metabolism and other cellular and biochemical functions or in other words, environmentally responsive genes (Li et al., 2016). The extended repertoire of almost all plant lectin genes in higher plants (Van Holle et al., 2017) suggests that these are multi-copy genes. However, it remains to be investigated whether there is a direct correlation between gene function

and gene duplicability, and how lectin genes have contributed to the adaptation of plants in a changing environment. In a recent study focusing on the immune response of *Arabidopsis* upon recognition of bacterial flagellin, the resilience of the plant immune system was explained by network buffering (Hillmer et al., 2017). Interactions among sectors of the network provide a basis for network buffering and can successfully compensate for the loss of a single component (Hillmer et al., 2017; Tyler, 2017). Since several plant lectins are reported to be involved in plant signaling (Xiang et al., 2011; Choi et al., 2014; Ranf et al., 2015; Couto and Zipfel, 2016; Balagué et al., 2017; Erwig et al., 2017; Xu et al., 2017), and given the strong expansion of lectin genes, partially as a result of polyploidization, homologous lectins are suggested to have subfunctionalized and could potentially facilitate network buffering in angiosperms.

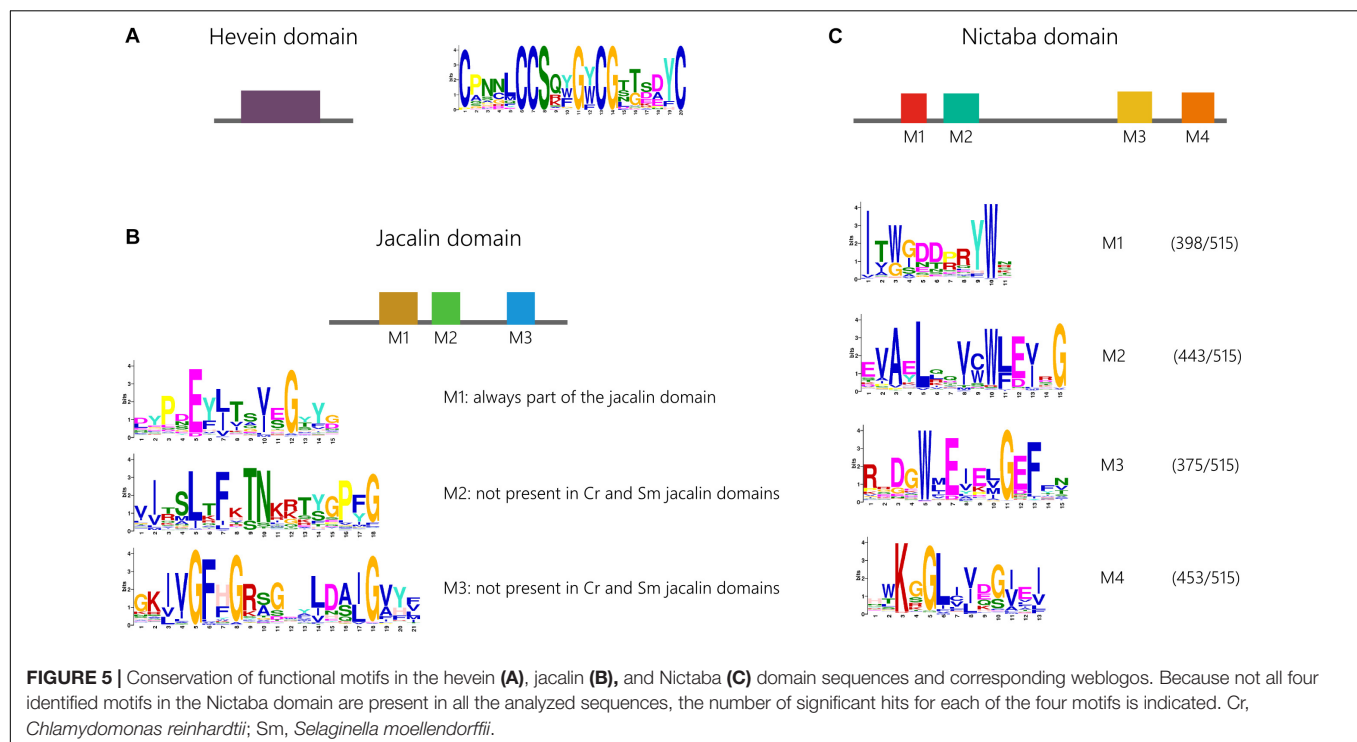
### Conserved Motifs in the Nictaba, Jacalin, and Hevein Lectin Domain Sequences

Motif analysis of the lectin domain sequences for the Nictaba, jacalin and hevein family was performed with MEME to analyze the retention of conserved motifs within the lectin domains across the different lineages (Figure 5). MEME analysis of the hevein domain sequences revealed only one significant motif, shared by sequences from all species under study (Supplementary Table 5). No significant differences were observed between the motif logo made for all sequences compared to the amino acid motif logo made based on the hevein domain sequences from *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max* and *Cucumis sativus*. Several cysteine and

glycine residues are highly conserved in this motif known to be important for the structure and folding of the hevein domain (Aboitiz et al., 2004). It can be concluded that this motif within the hevein domain is very conserved as it was already part of the hevein domain in both chlorophytes and Phragmoplastophyta.

Analysis of the jacalin domain sequences identified three different motifs, M1-M3 (Figure 5). The order of the three motifs was found to be highly conserved. Further analysis acknowledges motif M1 to be most retained. Moreover, this is the only significant motif that could be identified in jacalin domains from *Chlamydomonas reinhardtii* and *Selaginella moellendorffii*. Jacalin domain sequences from both mosses *Physcomitrella patens* and *Sphagnum fallax* do contain all three motifs, pointing to distinct evolutionary paths. This is again illustrated in the phylogenetic tree in Supplementary Figure 6 in which *Physcomitrella patens* and *Sphagnum fallax* sequences are grouped in separate branches. In gymnosperms and most angiosperms, all three motifs are present.

In the Nictaba domain, four significant motifs (M1-M4) were identified. Most of the sequences contain all four motifs in a specific order (Figure 5). However, there is a considerable number of sequences that only contain three or two motifs. The M2 and M4 motif are retained in 86% and 88% of all sequences, respectively. In contrast, M1 and M2 on the one hand, and M3 and M4 on the other hand, are often both present or absent in sequences that do not contain all four motifs. There is no strong correlation between the origin of the sequence (species) and the preservation of the four motifs. Nor are there significant differences in the sequences of the motif made for



all sequences, or for a subset of domain sequences representing the four model angiosperms. Remarkably, the M2 motif is absent in all Polypodiopsida sequences. Except for the M1 motif, all motifs were also retrieved in one or two *Chlamydomonas reinhardtii* sequences and the M4 motif is absent in *Chara braunii*. In the Nictaba domain sequences from the charophyte *Klebsormidium nitens*, all motifs are present with high confidence levels. These data suggest that the M1-M4 motifs originate from an ancestor of Viridiplantae, and that these motifs were not prone to substitution during further evolution.

Several amino acids that were designated to be crucial for the carbohydrate-binding activity of Nictaba, jacalin and hevein lectins, are part of the identified conserved motifs of the lectin domains (the tryptophan residues in M1 for the Nictaba domain; a leucine, tyrosine and aspartic acid residue in M3 of the jacalin domain and the serine and tyrosine residues in the motif identified in the hevein domain). Nevertheless, homologous plant lectins of one particular family can display different carbohydrate-binding specificities (Houlès Astoul et al., 2002; Fouquaert et al., 2009b; Fouquaert and Van Damme, 2012; Agostino et al., 2015). Since no clade-specific motifs were identified in the domain sequences of the Nictaba, jacalin and hevein domain; it is obvious that the conserved amino acids do not act as determinants of carbohydrate-binding specificity. Indeed, it has been reported that the carbohydrate-binding specificity of a lectin domain can change due to amino acid substitutions in loops or sequences which, upon folding of the polypeptide, are located in close vicinity of the binding site. Consequently, different carbohydrate-binding specificities between closely related lectins must result from other determinants that are not part of these motifs or amino acids within a motif at a position that displays sequence variability.

## CONCLUSION

To increase our understanding of the plant lectin families, we examined the origin of these protein families in the tree of life, with emphasis on the plant lineage. The widespread taxonomical distribution of some plant lectin domains was already described for the GNA, LysM, and ricin B lectin family, while the origin of the jacalin and Nictaba family was revisited. Taken together, our results suggest that different plant lectin families evolved in distinct ways. We documented variations of evolutionary paths at different levels, ranging from horizontal gene transfer, recombination of protein domains and discrepancy in gene loss and duplication events.

The evolution of lectins is characterized by expansion of the different lectin families from algae to higher plants, alongside with the diversification of lectins in terms of domain architecture and possibly functionality. Homologs of most lectin families are also present in extant representatives of charophytes and chlorophytes. Unmistakably, the magnitude of the plant lectin family in rhodophytes, glaucophytes and prasinophytes is far less than that observed in tracheophytes.

Only two groups of lectin motifs (LysM and ricin B) have been traced back to glaucophytes/rhodophytes. These are the most abundant plant lectin motifs in Bacteria, indicating that these plant lectin domains are most highly dispersed throughout the tree of life. Many of the essential plant features in land plants have their roots in charophyte algae (Leliaert et al., 2012; Umen, 2014; de Vries and Archibald, 2018; Nishiyama et al., 2018). Several lectin families have been detected in both charophyte algae and Streptophyta, indicating that many lectins originated before the evolution of land plants, and diversified later on. Regarding the domain architecture, an important number of the lectin sequences identified in sisters of vascular plants and streptophytes show resemblances to the domain architecture of animal lectins. It is clear that most of these sequences were not retained during diversification from algae to modern angiosperms. Other lectin domain architectures (e.g., F-box/Nictaba) arose from these ancestral lineages and are conserved in higher plants.

Most lectin sequences encode multi-domain proteins containing at least one lectin domain, suggesting that these proteins exert multiple biological activities. However, it remains challenging to predict the functionality of these lectins based on the domain sequences. Functional studies are needed to better understand their physiological roles. Although our knowledge of plant lectins has increased tremendously, a number of aspects on their evolutionary history remain incompletely understood. In the future, the availability of high quality chromosome-scale assemblies of more (plant) genomes will allow more detailed analyses (Rensing, 2017, 2018). It is apparent that the number of publications addressing the evolution of particular protein families in plants is increasing, and future research will without doubt enhance our understanding of this topic.

## AUTHOR CONTRIBUTIONS

SVH and EVD outlined and designed the study. SVH performed the research, analyzed the data, and prepared the manuscript. EVD conceived and supervised the study and critically revised the manuscript. All authors have read, revised, and approved the final version of the manuscript.

## FUNDING

This work was supported by the grants from the Fund for Scientific Research–Flanders (FWO Grant G006114N) and the Research Council of Ghent University (BOF15/GOA/005).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00036/full#supplementary-material>

## REFERENCES

- Aboitiz, N., Vila-Perelló, M., Groves, P., Asensio, J. L., Andreu, D., Cañada, F. J., et al. (2004). NMR and modeling studies of protein-carbohydrate interactions: synthesis, three-dimensional structure, and recognition properties of a minimum hevein domain with binding affinity for chitooligosaccharides. *Chembiochem* 5, 1245–1255. doi: 10.1002/cbic.200400025
- Agostino, M., Velkov, T., Dingjan, T., Williams, S. J., Yuriev, E., and Ramsland, P. A. (2015). The carbohydrate-binding promiscuity of *Euonymus europaeus* lectin is predicted to involve a single binding site. *Glycobiology* 25, 101–114. doi: 10.1093/glycob/cwu095
- Akcapinar, G. B., Kappel, L., Sezerman, O. U., and Seidl-Seiboth, V. (2015). Molecular diversity of LysM carbohydrate-binding motifs in fungi. *Curr. Genet.* 61, 103–113. doi: 10.1007/s00294-014-0471-9
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME suite: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Balagué, C., Gouget, A., Bouchez, O., Souriac, C., Haget, N., Bouter-Mercey, S., et al. (2017). The *Arabidopsis thaliana* lectin receptor kinase LecRK-I.9 is required for full resistance to *Pseudomonas syringae* and affects jasmonate signalling. *Mol. Plant Pathol.* 18, 937–948. doi: 10.1111/mpp.12457
- Bauters, L., Naalden, D., and Gheysen, G. (2017). The distribution of lectins across the phylum nematoda: a genome-wide search. *Int. J. Mol. Sci.* 18:91. doi: 10.3390/ijms18010091
- Bovi, M., Carrizo, M. E., Capaldi, S., Perduca, M., Chiarelli, L. R., Galliano, M., et al. (2011). Structure of a lectin with antitumoral properties in king bolete (*Boletus edulis*) mushrooms. *Glycobiology* 21, 1000–1009. doi: 10.1093/Glycob/Cwr012
- Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., et al. (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* 171, 287–304. doi: 10.1016/j.cell.2017.09.030
- Brawley, S. H., Blouin, N. A., Ficko-Blean, E., Wheeler, G. L., Lohr, M., Goodson, H. V., et al. (2017). Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *Proc. Natl. Acad. Sci. U.S.A.* 114, E6361–E6370. doi: 10.1073/pnas.1703088114
- Buist, G., Steen, A., Kok, J., and Kuipers, O. P. (2008). LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol. Microbiol.* 68, 838–847. doi: 10.1111/j.1365-2958.2008.06211.x
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Chater, C. C., Caine, R. S., Fleming, A. J., and Gray, J. E. (2017). Origins and evolution of stomatal development. *Plant Physiol.* 174, 624–638. doi: 10.1104/pp.17.00183
- Choi, J., Tanaka, K., Liang, Y., Cao, Y., Lee, S. Y., and Stacey, G. (2014). Extracellular ATP, a danger signal, is recognized by DORN1 in *Arabidopsis*. *Biochem. J.* 463, 429–437. doi: 10.1042/BJ20140666
- Couto, D., and Zipfel, C. (2016). Regulation of pattern recognition receptor signalling in plants. *Nat. Rev. Immunol.* 16, 537–552. doi: 10.1038/nri.2016.77
- Dang, L., Rougé, P., and Van Damme, E. J. M. (2017). Amaranthin-like proteins with aerolysin domains in plants. *Front. Plant Sci.* 8:1368. doi: 10.3389/fpls.2017.01368
- De Schutter, K., Tsaneva, M., Kulkarni, S. R., Rougé, P., Vandepoele, K., and Van Damme, E. J. M. (2017). Evolutionary relationships and expression analysis of EUL domain proteins in rice (*Oryza sativa*). *Rice* 10:26. doi: 10.1186/s12284-017-0164-3
- de Vries, J., and Archibald, J. M. (2018). Plant evolution: landmarks on the path to terrestrial life. *New Phytol.* 217, 1428–1434. doi: 10.1111/nph.14975
- Ekman, D., Björklund, Å. K., and Elofsson, A. (2007). Quantification of the elevated rate of domain rearrangements in Metazoa. *J. Mol. Biol.* 372, 1337–1348. doi: 10.1016/j.jmb.2007.06.022
- Ekman, D., Björklund, Å. K., Frey-Skött, J., and Elofsson, A. (2005). Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *J. Mol. Biol.* 348, 231–243. doi: 10.1016/j.jmb.2005.02.007
- Eme, L., Spang, A., Lombard, J., Stairs, C. W., and Ettema, T. J. G. (2017). Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* 15, 711–723. doi: 10.1038/nrmicro.2017.133
- Erwig, J., Ghareeb, H., Kopischke, M., Hacke, R., Matei, A., Petutschnig, E., et al. (2017). Chitin-induced and CHITIN ELICITOR RECEPTOR KINASE1 (CERK1) phosphorylation-dependent endocytosis of *Arabidopsis thaliana* LYSIN MOTIF-CONTAINING RECEPTOR-LIKE KINASE5 (LYK5). *New Phytol.* 215, 382–396. doi: 10.1111/nph.14592
- Forterre, P. (2015). The universal tree of life: an update. *Front. Microbiol.* 6:717. doi: 10.3389/fmicb.2015.00717
- Fouquaert, E., Peumans, W. J., Vandekerckhove, T. T. M., Ongenaert, M., and Van Damme, E. J. M. (2009a). Proteins with an *Euonymus* lectin-like domain are ubiquitous in Embryophyta. *BMC Plant Biol.* 9:136. doi: 10.1186/1471-2229-9-136
- Fouquaert, E., Smith, D. F., Peumans, W. J., Proost, P., Balzarini, J., Savvides, S. N., et al. (2009b). Related lectins from snowdrop and maize differ in their carbohydrate-binding specificity. *Biochem. Biophys. Res. Commun.* 380, 260–265. doi: 10.1016/j.bbrc.2009.01.048
- Fouquaert, E., and Van Damme, E. J. M. (2012). Promiscuity of the *Euonymus* carbohydrate-binding domain. *Biomolecules* 2, 415–434. doi: 10.3390/biom2040415
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, 78–86. doi: 10.1093/nar/gkr944
- Ham, J. H., Melanson, R. A., and Rush, M. C. (2011). *Burkholderia glumae*: next major pathogen of rice? *Mol. Plant Pathol.* 12, 329–339. doi: 10.1111/j.1364-3703.2010.00676.x
- Harrison, C. J. (2017). Development and genetics in the evolution of land plant body plans. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20150490. doi: 10.1098/rstb.2015.0490
- Hillmer, R. A., Tsuda, K., Rallapalli, G., Asai, S., Truman, W., Papke, M. D., et al. (2017). The highly buffered *Arabidopsis* immune signaling network conceals the functions of its components. *PLoS Genet.* 13:e1006639. doi: 10.1371/journal.pgen.1006639
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281
- Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., et al. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* 5:3978. doi: 10.1038/ncomms4978
- Houlès Astoul, C., Peumans, W. J., Van Damme, E. J., Barre, A., Bourne, Y., and Rougé, P. (2002). The size, shape and specificity of the sugar-binding site of the jacalin-related lectins is profoundly affected by the proteolytic cleavage of the subunits. *Biochem. J.* 367, 817–824. doi: 10.1042/bj20020856
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nat. Microbiol.* 1:16048. doi: 10.1038/nmicrobiol.2016.48
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., and Jermini, L. S. (2017). Modelfinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kanagawa, M., Liu, Y., Hanashima, S., Ikeda, A., Chai, W., Nakano, Y., et al. (2014). Structural basis for multiple sugar recognition of jacalin-related human ZG16p lectin. *J. Biol. Chem.* 289, 16954–16965. doi: 10.1074/jbc.M113.539114
- Katoh, K., Rozewicki, J., and Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* doi: 10.1093/bib/bbx108 [Epub ahead of print].
- Kubo, T., Abe, J., Saito, T., and Matsuda, Y. (2002). Genealogical relationships among laboratory strains of *Chlamydomonas reinhardtii* as inferred from matrix metalloprotease genes. *Curr. Genet.* 41, 115–122. doi: 10.1007/s00294-002-0284-0
- Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* 93, 515–533. doi: 10.1111/tpj.13801
- Lee, J., Park, J., Kim, S., Park, I., and Seo, Y. S. (2016). Differential regulation of toxoflavin production and its role in the enhanced virulence of *Burkholderia gladioli*. *Mol. Plant Pathol.* 17, 65–76. doi: 10.1111/mpp.12262



- Lehti-Shiu, M. D., and Shiu, S.-H. (2012). Diversity, classification and function of the plant protein kinase superfamily. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 2619–2639. doi: 10.1098/rstb.2012.0003
- Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F., et al. (2012). Phylogeny and molecular evolution of the green algae. *CRC Crit. Rev. Plant Sci.* 31, 1–46. doi: 10.1080/07352689.2011.615705
- Li, F., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M., et al. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* 4, 460–472. doi: 10.1038/s41477-018-0188-8
- Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y., and De Smet, R. (2016). Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell* 28, 326–344. doi: 10.1105/tpc.15.00877
- Maes, M., Huvenne, H., and Messens, E. (2009). *Brenneria salicis*, the bacterium causing watermark disease in willow, resides as an endophyte in wood. *Environ. Microbiol.* 11, 1453–1462. doi: 10.1111/j.1462-2920.2009.01873.x
- Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S. Y., et al. (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428, 653–657. doi: 10.1038/nature02398
- Mayer, S., Raulf, M. K., and Lepenies, B. (2017). C-type lectins: their network and roles in pathogen recognition and immunity. *Histochem. Cell Biol.* 147, 223–237. doi: 10.1007/s00418-016-1523-7
- Mitchell, A. L., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., et al. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* 43, 213–221. doi: 10.1093/nar/gku1243
- Naganuma, T., Hoshino, W., Shikanai, Y., Sato, R., Liu, K., Sato, S., et al. (2014). Novel matrix proteins of *Pteria penguin* pearl oyster shell nacre homologous to the jacalin-related  $\beta$ -prism fold lectins. *PLoS One* 9:e112326. doi: 10.1371/journal.pone.0112326
- Nagata, Y., Yamashita, M., Honda, H., Akabane, J., Uehara, K., Saito, A., et al. (2005). Characterization, occurrence, and molecular cloning of a lectin from *Grifola frondosa*: jacalin-related lectin of fungal origin. *Biosci. Biotechnol. Biochem.* 69, 2374–2380. doi: 10.1271/bbb.69.2374
- Nguyen, L., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K. K., et al. (2018). The chara genome: secondary complexity and implications for plant terrestrialization. *Cell* 174, 448–464. doi: 10.1016/j.cell.2018.06.033
- Panchy, N., Lehti-Shiu, M. D., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523
- Percudani, R., Montanini, B., and Ottonello, S. (2005). The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes. *Proteins* 60, 670–678. doi: 10.1002/prot.20543
- Peumans, W. J., Fouquaert, E., Jauneau, A., Rougé, P., Lannoo, N., Hamada, H., et al. (2007). The liverwort *Marchantia polymorpha* expresses orthologs of the fungal *Agaricus bisporus* agglutinin family. *Plant Physiol.* 144, 637–647. doi: 10.1104/pp.106.087437
- Price, D. C., Chan, C. X., Yoon, H. S., Yang, E. C., Qiu, H., Weber, A. P. M., et al. (2012). *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* 335, 843–847. doi: 10.1126/science.1213561
- Proost, S., Bel, M., Van, Vaneechoutte, D., Van De Peer, Y., Inzé, D., Mueller-Roeber, B., et al. (2015). PLAZA 3.0: An access point for plant comparative genomics. *Nucleic Acids Res.* 43, 974–981. doi: 10.1093/nar/gku986
- Puttick, M. N., Morris, J. L., Williams, T. A., Cox, C. J., Edwards, D., Kenrick, P., et al. (2018). The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr. Biol.* 28, 733–745. doi: 10.1016/j.cub.2018.01.063
- Ranf, S., Gisch, N., Schäffer, M., Illig, T., Westphal, L., Knirel, Y. A., et al. (2015). A lectin S-domain receptor kinase mediates lipopolysaccharide sensing in *Arabidopsis thaliana*. *Nat. Immunol.* 16, 426–433. doi: 10.1038/ni.3124
- Rensing, S. A. (2017). Why we need more non-seed plant models. *New Phytol.* 216, 355–360. doi: 10.1111/nph.14464
- Rensing, S. A. (2018). Great moments in evolution: the conquest of land by plants. *Curr. Opin. Plant Biol.* 42, 49–54. doi: 10.1016/j.pbi.2018.02.006
- Santamaría, M. E., Diaz-Mendoza, M., Diaz, I., and Martinez, M. (2014). Plant protein peptidase inhibitors: an evolutionary overview based on comparative genomics. *BMC Genomics* 15:812. doi: 10.1186/1471-2164-15-812
- Schallus, T., Jaeckh, C., Fehér, K., Palma, A. S., Liu, Y., Simpson, J. C., et al. (2008). Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation. *Mol. Biol. Cell* 19, 3404–3414. doi: 10.1091/mbc.e08-04-0354
- Soltis, P. S., Marchant, D. B., Van de Peer, Y., and Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125. doi: 10.1016/j.gde.2015.11.003
- Soltis, P. S., and Soltis, D. E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* 30, 159–165. doi: 10.1016/j.pbi.2016.03.015
- Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernet, B., and Durand, D. (2012). Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28, i409–i415. doi: 10.1093/bioinformatics/bts386
- Tyler, B. M. (2017). The fog of war: how network buffering protects plants' defense secrets from pathogens. *PLoS Genet.* 13:e1006713. doi: 10.1371/journal.pgen.1006713
- Umen, J. G. (2014). Green algae and the origins of multicellularity in the plant kingdom. *Cold Spring Harb. Perspect. Biol.* 6:a016170. doi: 10.1101/cshperspect.a016170
- Ura, H., Furuya, N., Iiyama, K., Hidaka, M., Tsuchiya, K., and Matsuyama, N. (2006). *Burkholderia gladioli* associated with symptoms of bacterial grain rot and leaf-sheath browning of rice plants. *J. Gen. Plant Pathol.* 72, 98–103. doi: 10.1007/s10327-005-0256-6
- Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., et al. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* 46, D1190–D1196. doi: 10.1093/nar/gkx1002
- Van Damme, E. J. M., Lannoo, N., and Peumans, W. J. (2008). Plant lectins. *Adv. Bot. Res.* 48, 107–209. doi: 10.1016/S0065-2296(08)00403-5
- Van Damme, E. J. M., Peumans, W. J., Barre, A., and Rougé, P. (1998). Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *CRC Crit. Rev. Plant Sci.* 17, 575–692. doi: 10.1080/07352689891304276
- Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424. doi: 10.1038/nrg.2017.26
- Van Holle, S., De Schutter, K., Eggermont, L., Tsaneva, M., Dang, L., and Van Damme, E. J. M. (2017). Comparative study of lectin domains in model species: new insights into evolutionary dynamics. *Int. J. Mol. Sci.* 18:1136. doi: 10.3390/ijms18061136
- Van Holle, S., and Van Damme, E. J. M. (2018). Signaling through plant lectins: modulation of plant immunity and beyond. *Biochem. Soc. Trans.* 36, 221–247. doi: 10.1042/BST20170371
- Van Hove, J., De Jaeger, G., De Winne, N., Guisez, Y., and Van Damme, E. J. M. (2015). The *Arabidopsis* lectin EULS3 is involved in stomatal closure. *Plant Sci.* 238, 312–322. doi: 10.1016/j.plantsci.2015.07.005
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4859–E4868. doi: 10.1073/pnas.1323926111
- Williams, T. A., Foster, P. G., Cox, C. J., and Embley, T. M. (2013). An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504, 231–236. doi: 10.1038/nature12779
- Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579. doi: 10.1073/pnas.87.12.4576
- Wong, J. E. M. M., Alsarraf, H. M. A. B., Kaspersen, J. D., Pedersen, J. S., Stougaard, J., Thirup, S., et al. (2014). Cooperative binding of LysM domains determines the carbohydrate affinity of a bacterial endopeptidase protein. *FEBS J.* 281, 1196–1208. doi: 10.1111/febs.12698
- Xia, X., You, M., Rao, X.-J., and Yu, X.-Q. (2018). Insect C-type lectins in innate immunity. *Dev. Comp. Immunol.* 83, 70–79. doi: 10.1016/j.dci.2017.11.020
- Xiang, Y., Song, M., Wei, Z., Tong, J., Zhang, L., Xiao, L., et al. (2011). A jacalin-related lectin-like gene in wheat is a component of the plant defence system. *J. Exp. Bot.* 62, 5471–5483. doi: 10.1093/jxb/err226

- Xing, S., Li, M., and Liu, P. (2013). Evolution of S-domain receptor-like kinases in land plants and origination of S-locus receptor kinases in Brassicaceae. *BMC Evol. Biol.* 13:69. doi: 10.1186/1471-2148-13-69
- Xu, J., Wang, G., Wang, J., Li, Y., Tian, L., Wang, X., et al. (2017). The lysin motif-containing proteins, Lyp1, Lyk7 and LysMe3, play important roles in chitin perception and defense against *Verticillium dahliae* in cotton. *BMC Plant Biol.* 17:148. doi: 10.1186/s12870-017-1096-1
- Young, J. M., and Park, D. C. (2007). Relationships of plant pathogenic enterobacteria based on partial *atpD*, *carA*, and *recA* as individual and concatenated nucleotide and peptide sequences. *Syst. Appl. Microbiol.* 30, 343–354. doi: 10.1016/j.syapm.2007.03.002
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Backstrom, D., Juzokaite, L., Vancaester, E., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358. doi: 10.1038/nature21031
- Zeng, F., and Harris, R. C. (2014). Epidermal growth factor, from gene organization to bedside. *Semin. Cell Dev. Biol.* 28, 2–11. doi: 10.1016/j.semcdb.2014.01.011
- Zhang, X.-C., Cannon, S. B., and Stacey, G. (2009). Evolutionary genomics of LysM genes in land plants. *BMC Evol. Biol.* 9:183. doi: 10.1186/1471-2148-9-183
- Zipfel, C., and Oldroyd, G. E. D. (2017). Plant signalling in symbiosis and immunity. *Nature* 543, 328–336. doi: 10.1038/nature22009

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Van Holle and Van Damme. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.